ERIM

ENVIRONMENTAL RESEARCH INSTITUTE OF MICHIGAN

P.O. Box 8618

Ann Arbor, Michigan 48107

Progress Report

for

NASA Contract NAS9-15476

ANALYSIS OF SCANNER DATA FOR CROP INVENTORIES

Program Manager

Quentin A. Holmes

Technical Manager

Robert Horvath

Task Coordinators

Richard J. Kauth

Richard C. Cicone

William A. Malila

Period Covered

15 November 1978 to 6 March 1979

ERIM Report Number 132400-20-P

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
132400-20-P		

4. Title and Subtitle	5. Report Date
Progress Report for November 15, 1978 - March 6, 1979 on the Analysis of Scanner Data for Crop Inventories	March 1979
	6. Performing Organization Code
7. Author(s) Co-Authors - R.J. Kauth, R.C. Cicone, W.A. Malila	8. Performing Organization Report No. 132400-20-P
9. Performing Organization Name and Address Environmental Research Institute of Michigan (ERIM) P.O. Box 8618 Ann Arbor, Michigan 48107	10. Work Unit No.
	11. Contract or Grant No. NAS9-15476
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Johnson Space Center Houston, Texas 77058	13. Type of Report and Period Covered Progress November 15, 1978 to March 6, 1979
	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

This report summarizes the planning and progress on the subject contract during the first Quarter.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Global crop inventory, sampling, haze correction, signature extension, Landsat, crop discrimination, photointerpretation	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
UNCLASSIFIED	UNCLASSIFIED		

*For sale by the National Technical Information Service, Springfield, Virginia 22161

JSC Form 1424 (Rev Nov 75)

NASA — J

PREFACE

The following report serves as the Quarterly report for Contract NAS9-15476 which is entitled, "Analysis of Scanner Data for Crop Inventories". This report describes the work carried out under that contract for the period 15 November 1978 through 6 March 1979.

Work on this contract is performed in the Infrared & Optics Division directed by Mr. Richard R. Legault. Dr. Quentin A. Holmes is the Project Manager for this contract and Mr. Robert Horvath is the Technical Manager.

This contract is part of a comprehensive and continuing program of research concerned with advancing the state-of-the-art in remote sensing of the environment from aircraft and satellites. The research is being carried out for NASA's Lyndon B. Johnson Space Center (JSC), Houston, Texas, by the Environmental Research Institute of Michigan (ERIM). The basic objective of this multidisciplinary program is to develop such information systems as practical tools which will provide planners and decision-makers extensive accurate information quickly and economically.

The principal focus during this reporting period was on planning. Hence, the Research Plan forms a substantial portion of the current report. Recent work on developing multicrop labeling aids in the form of color film products comprises the balance of this report. The materials which constitute this Quarterly Report were presented orally at the Supporting Research and Technology (S.R.&T.) Quarterly Review held at NASA/JSC on 7, 8, and 9 March 1979.

ERIM'S FY79 RESEARCH PLAN
Presented by Q. A. Holmes

The purpose of my speech today is to present to you the ERIM Research Plan for FY79. It is a bit of a misnomer to claim that this is a research plan for FY79 for there are elements in it which will require more than a year to be carried out. It is part and parcel of the nature of research that although one tries to plan ahead a foreseeable amount, not always are you given adequate vision to do so. There are some research problems that are well enough understood to be addressable within the time span of a single contract year, while there are other research objectives for which progress during a year of work serves as a good indication of how long will actually be required. As we laid this Research Plan out we thought to ourselves that it would be at most a two-year plan. We still have that in mind. Hopefully it will turn out that way, but if past experience is repeated, some of the items that I will be presenting today will still be beyond the state-of-the-art at the close of even the second year.

This is an outline of my presentation (VG-2). I will start by reminding you of a bit of background information, next I will summarize the major tasks and objectives of ERIM's S.R.&T. research and finally I will go over each of the subtasks by presenting some of the specific things that ERIM researchers will be doing.

It is a fact that although Landsat based remote sensing has been around now for more than seven years there are still some very basic things that we are not able to do. As remote sensing technologists, we feel very fortunate to have had Landsat. Indeed, the beauty of Landsat imagery is the beauty of the earth itself. Although we can feel a sense of pride in the tremendous progress which has been made by Landsat Investigators and by the technology thrust which LACIE has

ERIM RESEARCH PLAN FOR FY 1979

CONTRACT NAS9-15476

PRESENTED BY:

DR. QUENTIN HOLMES

PROGRAM MANAGER

MARCH 8, 1979

ERIM-1

OUTLINE

- BACKGROUND
- MAJOR TASKS AND OBJECTIVES
- SUBTASK STRUCTURE

ERIM-2

caused, the millenium obviously hasn't arrived yet. To some of you it will be a little bit strange to see the very same items on ERIM's list of research topics that have been there two, four, and even five years ago. In a sense that is ironic for I am sure that other NASA contractors before me have proposed researching many of those same topics and even accepted contracts in order to "solve this problem". However, this year we hope to do things a little differently.

The remote sensing community's experience in LACIE has left us all much wiser as to the magnitude of the problems which we face. Among other things, LACIE was the first large-scale program to apply the techniques that we have on a quasi-operational basis. In a sense those techniques were simply experimental techniques that were adopted for operational use because they were the best understood. For example, when LACIE adopted the "sum of likelihoods" classifier it did so because that was the only approach which had anything like a modest level of use behind it.

Many of us who were in remote sensing back when Landsat was first launched did not fully appreciate the difficulty of doing reliable economic information extraction from Landsat data in conjunction with collateral information over an area the size of a continent. Today we understand that this is a far more formidable task than we ever thought. In part the technical difficulties we encountered in LACIE stemmed from the fact that the testing that was done on the remote sensing techniques which were developed by individual principal investigators was usually limited to one or two data sets. Hence, when LACIE began we tended to implement techniques which performed well on some data sets, but whose performance on other data sets was less than optimal. Secondly, there is that harsh underlying fact that we scientists still do not know what it is that characterizes one scene and makes it similar to another scene. In spite of considerable effort by some of the best minds in our technology, such a characterization is something that still lies beyond today's state of the art. We do recognize that as Landsat gathers data over an area of the size of, say,

the southern great plains, that you tend to see a pretty broad ensemble of conditions. But what we don't fully appreciate is the fact that year-to-year variability is not well understood. The first three growing seasons worked by LACIE were each markedly different. In a sense, it was fortunate that LACIE ran for four years for it turned out that transition year 1978 was quite similar to one of the earlier years in terms of the climate and weather conditions which have a dramatic influence upon agronomic conditions.

There are only two major tasks and objectives in the very broad statement of work which ERIM has been given for this year and the following year (VC-3). These fundamental tasks are based upon two of the major limitations experienced in LACIE. That experience indicated that the most formidable problem the technology faces in a setting such as LACIE, where access to ground truth was forbidden, is to accurately label representative scene elements. In the case of LACIE Phase I and Phase II, one thinks of selecting and labeling training fields as wheat and non-wheat. Today the concern is centered around how to label dots as wheat and non-wheat. The first of ERIM's research tasks then is Objective Labeling Techniques Development. Here we shall seek to use remotely sensed data, in conjunction with collateral information, to allow a trained human interpreter, or a machine, to label selected scene elements such that these in turn can be used in a machine processing environment to estimate the proportion of wheat in an area. In a more general sense, we think of using selected scene elements which have been labeled in a production forecasting system to arrive at production estimates over an extensive region.

The second major task and objective of ERIM's Research is to make machine processing techniques worthwhile. There is a standing joke between us and our Technical Monitor who points out that ERIM was responsible for inventing the multispectral scanner and played a lead

MAJOR TASKS AND OBJECTIVES

TASK 1: OBJECTIVE LABELING TECHNIQUES DEVELOPMENT

OBJECTIVE: TO DEVELOP THE TECHNOLOGY TO ACCURATELY AND OBJECTIVELY LABEL SCENE ELEMENTS BASED ON REMOTELY SENSED AND COLLATERAL DATA. THE FOCUS THIS YEAR WILL BE MAINLY ON WHEAT.

TASK 2: MACHINE PROCESSING TECHNIQUES DEVELOPMENT

OBJECTIVE: TO CREATE AN ENVIRONMENT FOR EFFECTIVE LABELING AND FOR EFFICIENT USE OF LABELS IN A LARGE-SCALE CROP PRODUCTION FORECASTING SYSTEM.

ERIM-3

role in getting it declassified. "Now that the world has multispectral data", he keeps saying, "Don't you feel you also have the obligation to make machine processing worthwhile?" That statement itself is in need of some explanation. There have been great strides since the multispectral scanner was first brought out: both in terms of understanding what it is that the scanner signals respond to and in identifying the limitations of a classifier when applied over an extensive area. We have learned quite a bit about preprocessing data stabilization in terms of atmospheric corrections and developed ways to present the data in a stable or robust manner. But just as soon as one person or group creates a new way to process digital data to render it more robust, someone else takes these digital data and uses them to build a more stable film product (i.e., an analog display as it were for a human being to interpret). Thereby the analyst interpreters are helped and again the question arises, "when are you going to do something for us on the machine processing side? Those preprocessing ideas were to help the A.I.'s!" The truth is that, given the formidable task they face, analyst interpreters need all the help that they can get. The purpose of machine processing techniques development is two-fold; to create an environment for effective labeling, and for efficient use of those labeled entities in a large-scale crop production forecasting system.

In the LACIE and LACIE Transition programs, we have come to understand that one of the critical sources of errors in today's system stem from the errors of omission and commission that occur at labeling time. In one sense, those of us in the S.R.&T. community are accused of being irrelevant. Irrelevant because it is very easy to devise a sampling scheme, a classifier if you will, which allows you to take ground truth labels and render proportion estimates that are unbiased with respect to the labels. The truth of the matter is that any good classifier which has been properly designed and bias-corrected will produce estimate which are unbiased with respect to the source of labels. As long as researchers use the ground truth to do the labeling and the ground truth to do the

grading, they really haven't proved anything terribly important for a quasi-operational program. A far more stringent test is how well the procedure performs when the labels are provided by an analyst interpreter, using the kind of materials that would be available to him or her in a quasi-operational mode. Here the apologies and excuses begin. When the source of labeling is not the same as the source of grading, you can begin to tell apart the difference between classifiers. What we are seeking is a classifier which is robust, i.e., able to absorb a certain amount of labeling errors which creep in along the way as always happens in any large-scale system. Thus, machine processing techniques development is geared to understanding how to use the machine cleverly once you are given labels.

Today in LACIE, the machine is used primarily as a tallying device. The available Landsat acquisitions are examined along with supporting information and labels are assigned to type-one dots and type-two dots at the same time. The type-one dots are used to start a clustering run and the resulting clusters are then labeled. Next, statistics from those labeled clusters are used to train a classifier. Technically speaking, you would do better by not using the type one dots for labeling the clusters, but instead allowed the clustering (and the classifying if you insist upon having a classifier in there) to segment the scene into strata which were then tallied according to the percentage of wheat and non-wheat dots they contained. Although that's about as good as you can do today, our hope is to go significantly beyond this through the research we will be doing under this S.R.&T. contract.

Now let us look at the summary of tasks and subtasks (VG-4). Under Objective Labeling Techniques, we being with feature definition studies. There exist two very different approaches or schools of thought on feature definition and members of these schools seldom talk to each other. One approach is taken by the school of remote sensing technologists who are concerned about the kind of signals that these sensors

SUMMARY OF TASKS AND SUBTASKS

TASK 1: OBJECTIVE LABELING TECHNIQUES DEVELOPMENT

- SUBTASK 1.1: FEATURE DEFINITION STUDIES
- SUBTASK 1.2: SIGNATURE CHARACTERIZATION STUDIES
- SUBTASK 1.3: LABELING TECHNIQUE DEVELOPMENT
- SUBTASK 1.4: DATA BASE DEVELOPMENT

TASK 2: MACHINE PROCESSING TECHNIQUES DEVELOPMENT

- SUBTASK 2.1: DATA NORMALIZATION STUDIES
- SUBTASK 2.2: BIAS AND VARIANCE REDUCTION TECHNIQUES
- SUBTASK 2.3: TEST AND EVALUATION
- SUBTASK 2.4: SYSTEM ERROR MODEL DEVELOPMENT
- SUBTASK 2.5: ADVANCED TECHNOLOGY STUDIES

ERIM -4

we have sent into space, and also use from aircraft, generate. People with access to a terminal or a small mini-computer get some remotely sensed digital data, and they become convinced that they are going to change the world; to solve things. Sometimes members of this school do pretty wee considering that they completely ignore a whole realm of scientific endeavor by researchers who study plants and plant physiology and try to understand and characterize, as best they can, how it is that vegetation grows, how it changes, and how it dies. What is needed more than anything else is some form of tie point between these two technologies so that technologists can relate remotely sensed signals back to something meaningful in the world of the plan physiologists. For only then can remote sensing begin to take advantage of the insight which they have accumulated about what it is that a plant is doing in response to stress, in response to too much rain, in response to very, very severe weather, in response to a limited amount of drought.

Our Feature Definition studies (VG-5) are characterized by trying to com up with ways to relate the physiology of a plant to the observable spectral characteristics as seen bv a satellite like Landsat. In a sense one thinks of modeling here. Modeling as we say "from seed to satellite".

The second ERIM subtask, entitled Signature Characterization Studies (VG-6) is aimed at taking the results of the Feature Definition Studies and using them to characterize the crop of interest. This year the crops that we will be concentrating on are winter wheat and spring wheat, with some minor attention paid to corn and soybeans. The intent is to do this work in such a way it will be of a more fundamental or basic nature so that it will be useful as a foundation for other crops at a later time. The presumption made at the outset in Signature Characterization Studies is that the proper setting for modelling crop signatures is a multitemporal environment with a sampling matched to the growing season

TASK 1 ACTIVITIES

SUBTASK 1.1: FEATURE DEFINITION STUDIES

OBJECTIVE: TO ESTABLISH QUALITATIVE AND QUANTITATIVE RELATIONSHIPS BETWEEN (A) AGRONOMIC AND PHYSIOLOGICAL DESCRIPTORS OF CROP STATE/ CROP DEVELOPMENT, AND (B) SPECTRAL-TEMPORAL FEATURES THAT HAVE POTENTIAL USE AS CROP DISCRIMINATORS.

PLANNED ACTIVITIES:

- STUDY STRUCTURE OF SPECTRAL DATA.
- CHARACTERIZE WHEAT SPECTRAL PHENOLOGY.
- EXAMINE AI PERFORMANCE PATTERNS.

TASK 1 ACTIVITIES (CONT.)

SUBTASK 1.2: SIGNATURE CHARACTERIZATION STUDIES

<u>OBJECTIVE</u>: TO INVESTIGATE METHODS OF CHARACTERIZING CROP SIGNATURES IN GENERAL, AND TO CHARACTERIZE THE MAJOR DISCRIMINATIVE FEATURES OF WHEAT IN PARTICULAR.

<u>PLANNED ACTIVITIES</u>:

- DEVELOP GENERALIZED SIGNATURE DESCRIPTION.
- INVESTIGATE STABILIZING TRANSFORMATIONS.

of the crop. At ERIM we don't think of ourselves as being constrained to the observables in Landsat but rather we believe that in order to truly characterize signatures one has to consider multiple acquisitions together with their collateral data.

I guess if I have a reservation about this it stems from the fact that I vividly remember more than two phases of LACIE in which some very excellent men spent the majority of their waking hours trying to do signature extension. As NASA and the S.R.&T. community marched down this road with some of us clustering while others tried preprocessing or different spectra matching techniques, it took a long time before we recognized that this concept was a blind alley. A blind alley not because there was no end to it, but a blind alley because no one got clever enough technically to make it work. As you recall, when LACIE first began, there was to be some 4,800 sample segments drawn from seven major wheat producing countries of the world. The idea was that about 1/5th of these 4800 5x6 nautical mile sample segments would be worked by analysts in the sense of labeling selected fields as to whether they were wheat or non-wheat. Then the signatures from these training segments would be used to process as many as five other segments in the same strata. That would have made it very cost-effective in terms of the amount of labeling required. Later on, it was felt that perhaps two or three training segments, in conjunction with each other, were needed to get an adequate statistical representation of a recognition segment. It was still believed that by labeling a small fraction of the segments, one could process all the segments in an entire strata. As you all know that didn't prove out in LACIE either. That didn't prove out because we aren't able to adequately characterize signatures today. Given an arbitrary sample segment, and an analyst interpreter to make the very necessary match between such incommensurate items as historical data, information about whether the current year is early or late in terms of getting a crop started, whether it is an abnormally

warm year or an abnormally wet year, and apply these to the task of interpreting the available Landsat acquisitions, then we can do a pretty fair job. But, unless we have a human being involved, unless he can see the segment, we have not yet been able to do proportion estimation very well.

Labeling Technique Development (VG-7) is the third subtask under the Objectives Labeling Techniques Development. Here we speak about drawing upon the outcome of both Subtask 1.1 and Subtask 1.2. Two major thrusts come to the fore. One seeks to extend the work that has been done to date on the LIST technique. The approach which will be taken is similar to one used by medical doctors wherein they give a patient a battery of 100 seemingly innocuous questions and the patients responses in turn are then subjected to statistical analysis to derive inferences as to the health of the patient. In a sense the analysts are asked to fill out a questionnaire on specific dots. A questionnaire which is couched in terms which are reasonable to ask an analyst interpreter to answer. Then, using correlation techniques, these responses are related to whether or not this particular dot ought to be labeled as the crop of interest, say wheat, or some other crop. In principal, this approach would allow some of the human biases that are involved to be overcome, by decoupling labeling from the direct influence of the analyst interpreters. This will be done while still keeping intact the analyst's all important role as the integrating point for information which is derived from noncommensurate sources.

There is a second aspect of Subtask 1.3. In Procedure M, the Multicrop Inventorying Procedure that we spoke of last November, there was a need to separate spring wheat from other spring small grains. We began Procedure M with the assumption that the LACIE analysts already have the ability to adequately separate spring small grains from non-small grains. The procedure developed was based upon the temporal track in Tasselled Cap space of a given picture element. This procedure

TASK 1 ACTIVITIES (CONT.)

SUBTASK 1.3: LABELING TECHNIQUES DEVELOPMENT

OBJECTIVE: TO PROVIDE SPECIFIC LABELING AIDS AND LABELING PROCEDURES, BASED ON THE IMPROVED UNDERSTANDING OF SIGNATURES AND DISCRIMINATIVE FEATURES GAINED IN SUBTASKS 1.1 AND 1.2.

TWO ACTIVITIES IN PROGRESS:

- IMPROVED ANALYST DISPLAY PRODUCTS
  - DETAILED PRESENTATION TO FOLLOW THIS OVERVIEW PRESENTATION
- IMPROVED MACHINE LABELING ALGORITHM FOR DISCRIMINATING SPRING WHEAT FROM OTHER SPRING SMALL GRAINS

PLANNED ACTIVITY:

- OBJECTIVE LABELING PROCEDURE FOR WHEAT

included a crop calendar shift on a pixel-by-pixel basis. It also included an automatic labeler which was built based upon LACIE field measurement data and modelling. This labeler, or decision maker, took pixels which had previously identified as spring small grains and divided them into two bins; spring wheat and other spring small grains. We plan to extend this labeling technique. You may recall the experiment we did whose results were reported on last December. Those results showed that the procedure performed quite well along the Red River Valley in South Dakota where the initial development was done. Areas where the labeler didn't perform well were later found to be characterized as areas where a severe drought had occurred. Although not surprising in retrospect, it certainly was surprising to us at the time, when the spring wheat signatures along the Red River Valley (which had an abundance of rain) looked like the non-wheat signatures in the regions to the west. The extension of the labeler which we are seeking would not make this mistake.

The other major subtasks that we have come under the category of Machine Processing Techniques Development. The first subtask under that category is Data Normalization (VG-9). This includes establishing a linear transformation for application to Landsat-3 so that existing techniques which are dependent upon the Landsat-2 calibration, (for example, the XSTAR atmospheric correction procedure, and also the parameters that we currently use in the clustering algorithms) can be taken over unchanged or transferred to the Landsat-3 data that we will be receiving from JSC. We are currently in the process of developing a Landsat-3 to a Landsat-2 transformation based upon sample segments that we have received from JSC.

The second subtask under Machine Processing Techniques Development is devising processing protocols which results in Reduced Bias and Variance (VG-10). In a sense we regard the role of a classifier in the large area inventory setting as inseparable from the sampling protocol. The

TASK 1 ACTIVITIES (CONT.)

SUBTASK 1.4: DATA BASE DEVELOPMENT

<u>OBJECTIVE</u>: TO CREATE AND MAINTAIN A DATA BASE THAT CAN BE ANALYZED AND USED IN THE DEVELOPMENT OF LABELING TECHNIQUES. IT WILL BUILD ON DATA BASE ESTABLISHED AT JSC AND ELSEWHERE.

<u>PRIMARY SOURCES</u>:

- ITS AND BLIND-SITE DATA.
- DATA BASE DEVELOPED FOR YIELD RESEARCH.
- LACIE FIELD MEASUREMENT DATA.

ERIM-8

TASK 2 ACTIVITIES

SUBTASK 2.1: DATA NORMALIZATION STUDIES

OBJECTIVES:

- TO ESTABLISH A TRANSFORMATION TO BE APPLIED TO LANDSAT 3 SO THAT EXISTING SCREENING AND ATMOSPHERE HAZE CORRECTION ALGORITHMS (DEVELOPED FOR LANDSAT 2) CAN BE UTILIZED ON LANDSAT 3 DATA.
- TO DETERMINE WHETHER OR NOT LANDSAT 2 CALIBRATION HAS DRIFTED SINCE JANUARY, 1978, AND, IF SO, ESTABLISH A CORRECTION ALGORITHM FOR IT.

ACTIVITIES IN PROGRESS:

- DEVELOPMENT OF LANDSAT 3 → LANDSAT 2 TRANSFORMATION.

TASK 2 ACTIVITIES (CONT.)

SUBTASK 2.2: BIAS AND VARIANCE REDUCTION TECHNIQUES

OBJECTIVES:

- TO IMPROVE SPECTRAL STRATIFICATION TO REDUCE SAMPLING VARIANCE.
- TO REDUCE BIAS CAUSED BY NOT SAMPLING THE "SMALL-FIELD STRATUM" IN PROCEDURE M.

ACTIVITIES IN PROGRESS:

- STUDY OF THE TOLERANCE-BLOCK APPROACH TO SPECTRAL STRATIFICATION.
- STRATIFICATION USING DERIVED SPECTRAL FEATURES MORE CLOSELY RELATED TO AGRONOMIC FEATURES.
- STUDY OF BIAS REDUCTION TECHNIQUES.

overall task is to use as many samples as you can afford to label in a very cost-effective manner. Hence, the stratification and the purity of those strata is the key to a high reduction in variance. If you have pure strata then a very small number of labeled samples would be adequate. The first objective here then is to learn how to produce improved spectral stratifications in order to reduce sampling variance.

Secondly, there is an entire stratum that gives everyone difficulty. That stratum is the one which is formed by boundary pixels. In Procedure M this stratum contains the small fields; for that procedure includes spectral/spatial clustering to produce blobs, then the outermost pixel around the perimeter of each blob is stripped away, and the blob interior is what is used as a labeling target. However, if after this stripping operation there are no interior pixels left in a blob, then there is no target to label. Thus, we seek to reduce the variance which we currently have by virtue of the fact that we don't sample this strata. We hope to use the results of the work in Task 1.1 to look at the greatest limitation of multispectral stratification. Our intent is to use derived spectral features which can be closely related to agronomic features. Here again we don't think of the actual Landsat signal values, but rather we speak about a trajectory in brightness-greenness-yellowness space during time and using that mathematical curve to characterize observable agronomic features so that we can understand physically what is going on. Finally we hope to study bias reduction techniques.

Under the Machine Processing Techniques Development we have come to understand that one of the major things that keeps you honest is to do trial data processing in a test and evaluation mode (VG-11). LACIE has contributed significantly to remote sensing. Not only was LACIE the first large scale project to use multitemporal data routinely for labeling and at times, for classification, but also it has made available the blind test sites and the associated ground truth, for Phase I, Phase II, and Phase III, and, soon we hope, for LACIE Transition. Collectively these

TASK 2 ACTIVITIES (CONT.)

SUBTASK 2.3: TEST AND EVALUATION

OBJECTIVES:

- TO MAINTAIN A CLEAR INDICATION OF WHICH PORTIONS OF THE OVERALL SYSTEM ARE MOST IN NEED OF IMPROVEMENT.
- TO DETERMINE THE CAUSES AND SOURCES OF ERROR.
- TO IDENTIFY THE NATURE OF PROBLEMS ASSOCIATED WITH EXPANDING LACIE-LIKE TECHNOLOGY TO NEW CROP TYPE AND REGION ENVIRONMENTS.

ACTIVITIES IN PROGRESS:

- TEST OF PROCEDURE M ON 16 TY 1978 SPRING WHEAT SEGMENTS.
- TRANSFER OF PROCEDURE M CAPABILITY TO THE PURDUE/LARS FACILITY TO PERMIT JSC TO CONDUCT INDEPENDENT TESTS AND EVALUATIONS OF PROCEDURE M.

PLANNED ACTIVITY:

- EXPLORATORY TESTS ON TY 1978 CORN AND SOYBEANS SEGMENTS.

ERIM-11

represent a substantial sample over space, the distribution of the United States as it were, and over time, both throughout the growing season, and from one growing season to the next. Given access to these data with the associated ground truth there is hope that researchers will be able to develop procedures which are robust enough to be valid over a broad area and over more than just a single year or growing season. The objective of Test and Evaluation is to maintain a clear indication of which portions of the overall system are most in need of improvement. We begin with the notion of using Procedure M as a test bed because the pieces are in place. We will use Procedure M then as a trial test bed, a control as it were, in which to embed the ideas that I have talked about earlier.

We speak about determining the causes and sources of errors and to identify the needs for solutions to the problems associated with expanding the LACIE technology to additional crops, and to different regions. We currently have in process a Procedure M experiment on 18 transition year (1978 growing seasons) spring wheat segments. Here, for the first time, an S.R.&T. group will be examining a procedure using analyst labels as inputs. Basically we are using JSC/LACIE interpreters to label the interiors of blobs that are created through the Procedure M environment. Those labels are then used in a very substantive series of experiments. (1) Three different analysts labeled each segment, and the experiment design has been laid out in such a manner that we will be able to ascertain the performance of each AI in labeling. In the sense that we will be able to compare analyst assigned labels with the ground truth that will become available later, the experiment will measure how well analysts label. This experiment will enable us to understand what it is that an AI can do well and what things are not realistic or unreasonable to ask him to do with a high level of accuracy. (2) The analyst labeling is being done exhaustively, that is, all of the blobs which have interiors are being labeled. The individual experiments themselves will be repeated more

than 50 times so that not only a proportion estimate but also ensemble means and standard deviations will be computed for each of these 18 sample segments. This will allow us to check the bias and, at the same time, measure how broad the variation is as one draws different numbers of labeling targets and varies the number of strata that are used. Although, as currently configured, P-1 uses two strata, Procedure M allows the use of 1, 2, 40, or even as many as 400 strata and that too is a parameter which will be examined. We are in the process of transferring Procedure M to the LARS Computer Facility because that is the computer that EOD runs on. We do this in order to make it available to you for further tests and evaluation. In one sense that is a drain on our resources because most of my researchers are so accustomed to working on our Amdahl 470 V6B. When that transfer is completed it will permit JSC and their in-house contractors to conduct independent test and evaluation of Procedure M. We are also planning to do some exploratory tests on some transition year sample segments (i.e., 1978 growing season corn and soybean segments) that we will be receiving.

The fourth subtask under the Machine Processing Techniques Development is System Error Model Development (VG-12). Again and again system level analyses confirm that no one has resources enough to be able to anticipate all the variability, all the conditions, that one will encounter in a quasi-operational system. A system error model provides two critical items. First, it allows you to identify which are the areas where one could gain appreciably if better techniques were available and, second, it allows you to map known conditions through the system to see what their effect is. We hope to develop a convenient representation of error sources and error propagation in Procedure M. As many of you know this procedure is somewhat akin to Procedure 1 in that it is responsive to the notion of trying to identify the sources of error. That was one of the major reasons for going

TASK 2 ACTIVITIES (CONT.)

SUBTASK 2.4: SYSTEM ERROR MODEL DEVELOPMENT

OBJECTIVES:

- TO DEVELOP A CONVENIENT REPRESENTATION OF ERROR SOURCES AND ERROR PROPAGATION IN PROCEDURE M.
- TO EXTEND THE MODEL TO MORE GENERALIZED SYSTEMS.

ERIM -12

to Procedure 1. We hope to extend the model to more generalized systems, for there are other systems beyond a Procedure 1 or a Procedure M. Systems in which the sampling is a more integral part of the stratification than the choice of an arbitrary allocation of 5x6 mile sample segments laid out, as it were, in advance.

The fifth subtask in Machine Processing Techniques Development is Advanced Technology Studies which are geared more toward the future (VG-13). The objective here is to generate viable technical options for future large-scale agricultural forecasting systems. We plan to examine a few selected study topics which are typical of the 1980-1985 time-frame and then go through the steps of generalized problem solving to identify key issues and begin work on resolving them. In particular we have begun to think about how is it that a machine should be used to extend label information to a large area when you're not constrained to pre-allocated sample segments.

You may recall that in the Secretary of Agriculture's Initiative, which is the key requirements document for the AGRISTARS program, the very first objective is on early warning systems. Yet, the technology that we have been developing in LACIE has been geared more toward a crop production forecasting system which is the second objective of AGRISTARS. Under the advanced technology study we plan to look at the technological basis for developing early warning systems. In a sense then the work that we'll be doing at ERIM during this coming year is intended to take off from exactly where we have arrived in LACIE--to build upon existing technology, to use available data to do test and evaluation, a moving forward of the technology in an orderly way from where we are today. This will include some very basic research aimed at addressing the issues which have gone inadequately answered since multispectral remote sensing first became of interest. It is rather naive of me to stand here and promise significant advances in all of these areas but our researchers aim to give it their best. It is our belief that machine

TASK 2 ACTIVITIES (CONT.)

SUBTASK 2.5: ADVANCED TECHNOLOGY STUDIES

OBJECTIVE: TO GENERATE VIABLE TECHNICAL OPTIONS FOR FUTURE LARGE-SCALE AGRICULTURAL FORECASTING SYSTEMS.

APPROACH:

- SELECT A FEW STUDY TOPICS WHICH ARE TYPICAL OF THE 1980-85 TIME-FRAME.
- GO THROUGH THE STEPS OF GENERALIZED PROBLEM SOLVING.
- IDENTIFY KEY TECHNICAL ISSUES AND QUESTIONS WHICH HAVE POTENTIAL TO BE RESOLVED.
- BEGIN WORK ON RESOLVING THE KEY QUESTIONS.

PLANNED FOCI:

- GENERAL CONSIDERATION OF THE USE OF MACHINE PROCESSING TO EXTEND LABEL INFORMATION TO A LARGE AREA.
- TECHNOLOGICAL BASIS FOR "EARLY WARNING" SYSTEMS.

ERIM-13

processing can be made worthwhile and that the difficulty currently associated with labeling arbitrary portions of scene elements, can be overcome by labeling more reliably, more robustly, more efficiently than it is done today. Only when that is done will this technology begin to reach the potential that all of us believe it is capable of providing; namely, providing cost-effective, reliable production estimates for an extensive area of the world.

ADVANCED FILM PRODUCT DEVELOPMENT

(R. J. Balon)

OBJECTIVE

The objective of this effort is to gain understanding and colorimetric control of the process of image production so that (1) visual color resolutions of data features are appropriately represented, (2) independent data features drive independent psychophysical dimensions of color, and (3) color meaning and interpretation are stable from image to image.

APPROACH

We have abandoned the concept, which has heretofore been standard, of making imagery by independently driving the three PFC primaries. Advanced film products will have the flexibility of directed color mapping. We map data vectors directly into a color space, i.e., a color specification and spacing system, which is expressed as a mathematical transformation on the PFC color primaries. The science of colorimetrics provides us a color space with dimensions of color specification and Euclidean spacing of colors in harmony with the established psychophysical properties of human color perception. We use this space as a screen for projection of data features. Properties of the color space make it possible to align independent data features with independent psychophysical dimensions of color and to control the visual color resolution of the features. Using a fixed map of data to color stabilizes color meaning and interpretation from scene to scene.

HIGHLIGHTS OF PROGRESS TO DATE

A model of the color production characteristics of the PFC has been formed [1]. The L*, a*, b* color space, which is the CIE 1976 standard

[1] Richard D. Juday, "Colorimetric Principles As Applied to Multichannel Imagery," National Aeronautics and Space Administration/Johnson Space Center, July 1978.

PRECEDING PAGE BLANK NOT FILMED

uniform color space, has been used up to the present as the model for properties of human color perception. Analysis has been completed on the current, operational method of producing LACIE imagery [2]. Product One imagery was found to: (1) be generally unstable in color meaning between scenes; (2) perceptually distort data variation in a characteristics manner; and (3) ignore potentially useful data variation, by virtue of dropping Channel 3. The Kraus product shares these drawbacks to a lesser degree, with the note that scene-to-scene distortion is primarily in brightness of colors rather than hue which reduces their impact on interpretation. A preliminary approach to uniform color space imagery has been developed. Research model computer software has been written to implement this approach and a limited amount of imagery has been produced.

PROGRESS IN THIS QUARTER

Progress was made this quarter toward finalizing the approach to making uniform color space imagery from Tasselled Cap transform data features with scene-to-scene haze correction. Figures 1 through 13 detail the goals of this work along with the specific technical design items put forth to achieve them. This design was used to make imagery for sample segment 1663. The results encourage us to believe uniform color space imagery can be made with a fixed mapping while providing adequate color resolution throughout the growing season.

Effort has been placed into obtaining actual film displays of the PFC color gun cube and L*, a*, b* uniform color space. Programs have been written to sample these color spaces with continuous slices and with discrete color "chips". These images of color space will be used: (1) to evaluate properties of the spaces (perceptual uniformity, "flatness" of constant lightness levels, and how the color space axes relate to the psychophysical dimensions of color); and (2) to fine tailor the placement

[2] R. J. Balon and R. C. Cicone, Technical Memorandum "Uniform Color Space Analysis of LACIE Image Products," ERIM No. 132400-10-R, May 1979.

of the data into color. Such color space slices will eventually also be the color keys for imagery made by directed color mappings.

OUTLOOK

The final design considerations are being finished for film products featuring Tasselled Cap data features, a standardized color mapping, and L*, a*, b* color control. It is expected that a product suitable for serious examination and testing will be ready about July. Longer term developments will depend upon gaining further knowledge about the data structure and the significance of features extracted from the data. The image research performed thus far is readily adaptable to use of different data features. In particular, we anticipate looking at variables derived from multitemporal data.

FIGURE 1. GOALS

- CAPTURE DATA VARIATION WITHOUT LOSS AND WITHOUT DISTORTION
- STANDARDIZE CORRESPONDENCE OF COLOR TO GROUND REFLECTANCES
- MATCH INDEPENDENT DIMENSIONS OF DATA VARIATION WITH INDEPENDENT PERCEPTUAL DIMENSIONS OF COLOR

FIGURE 2. STANDARDIZATION OF REPRESENTATION

- STANDARDIZE
  - A GIVEN SET OF GROUND REFLECTANCES SHOULD ALWAYS MAP TO THE SAME COLOR ON THE IMAGE PRODUCT

MOTIVATION

- IN THE PRESENCE OF VARIABLE COLOR MAPPINGS, GOOD COLOR KEYS CANNOT BE PROVIDED AND INTUITIVE PATTERN RECOGNITION IS CONFUSED
- COLOR IMAGES TAILORED TO SCENE STATISTICS CAN AND HAVE LED TO MISINTERPRETATION OF AREAS WITH DIFFERENT CROP PROPORTIONS AND AGRICULTURAL BALANCES
- SUCH IMAGES CAN AND HAVE CONFUSED CROP SIGNATURES IN MULTITEMPORAL IMAGERY

ERIM

FIGURE 3. APPROACH TO STANDARDIZATION

- STANDARDIZE DATA AGAINST EXTERNAL EFFECTS
  - SUN ANGLE CORRECTION
  - SPATIALLY VARYING HAZE CORRECTION ($XSTAR_{SV}$)
- STANDARDIZE CORRESPONDENCE OF COLOR TO DATA
  - USE A SINGLE, FIXED TRANSFORMATION FROM DATA TO COLOR FOR ALL SCENES AND ALL TIMES

FIGURE 4. PRESERVE POTENTIAL INFORMATION

- CAPTURE DATA VARIATION WITHOUT LOSS

MOTIVATION

- THE LABELING TASK NEEDS ALL AVAILABLE INFORMATION FROM THE SCANNER DATA

FIGURE 5. APPROACH TO PRESERVING INFORMATION

- SPAN THE DIMENSIONALITY OF THE DATA VARIATION

DIMENSIONALITY OF DATA	DIMENSIONALITY OF COLOR PERCEPTION
4 SCANNER BANDS	3 BRIGHTNESS, HUE, SATURATION

PROBLEM: REDUCE DIMENSIONALITY WITH MINIMAL INFORMATION LOSS

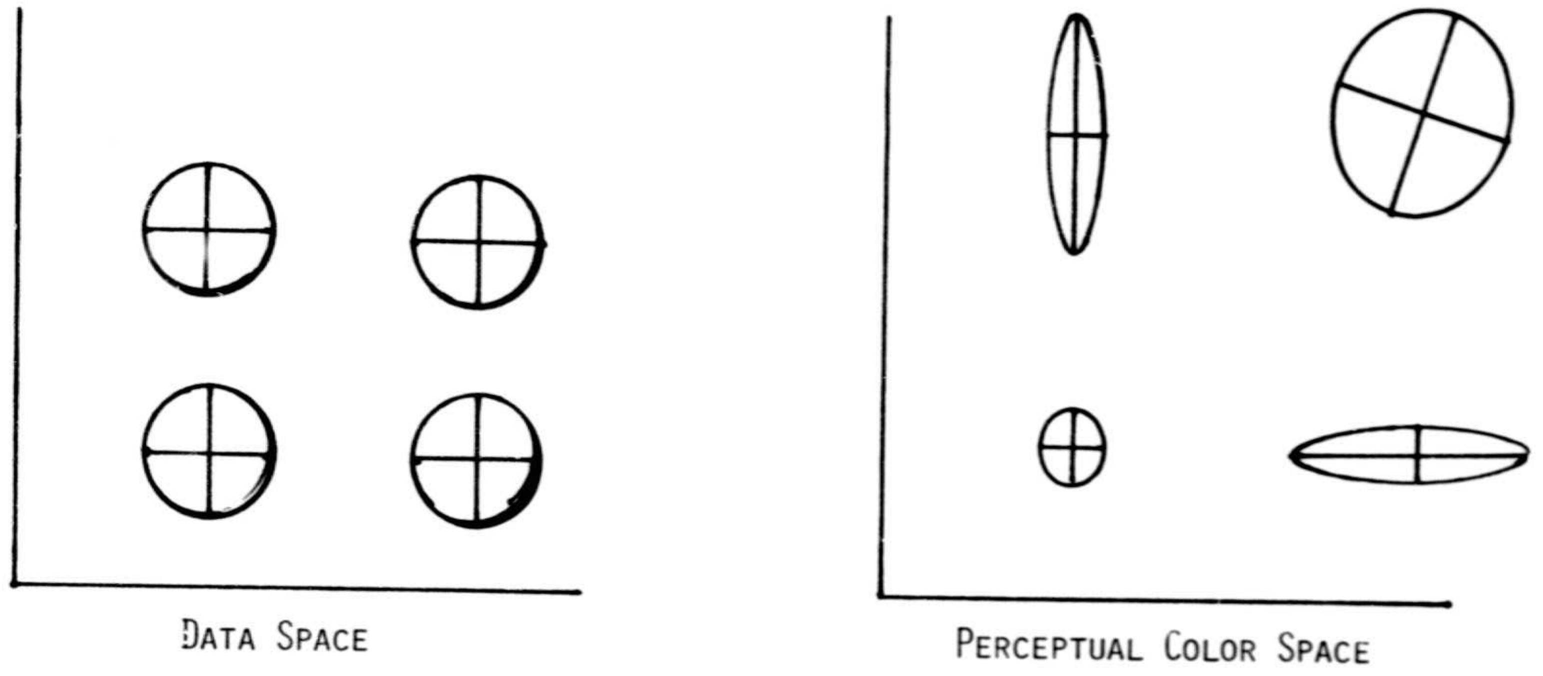
4 LANDSAT BANDS ⟶ LINEAR TASSELED-CAP TRANSFORMATION ⟶ TASSELED-CAP BRIGHTNESS, GREENNESS, YELLOWNESS, NONESUCH

FIRST THREE COMPONENTS SPAN 99% OF VARIATION

ERIM

FIGURE 6. CONTROL VISUAL IMPACT

• Capture Data Variation Without Distortion



Motivation

Gain Control Over Visual Impact of Spectral Features

ERIM

FIGURE 7. APPROACH TO CONTROL OF VISUAL IMPACT

- PLACE THE DATA DIRECTLY INTO THE COLOR ARRANGEMENT OF A PERCEPTUALLY UNIFORM COLOR SPACE

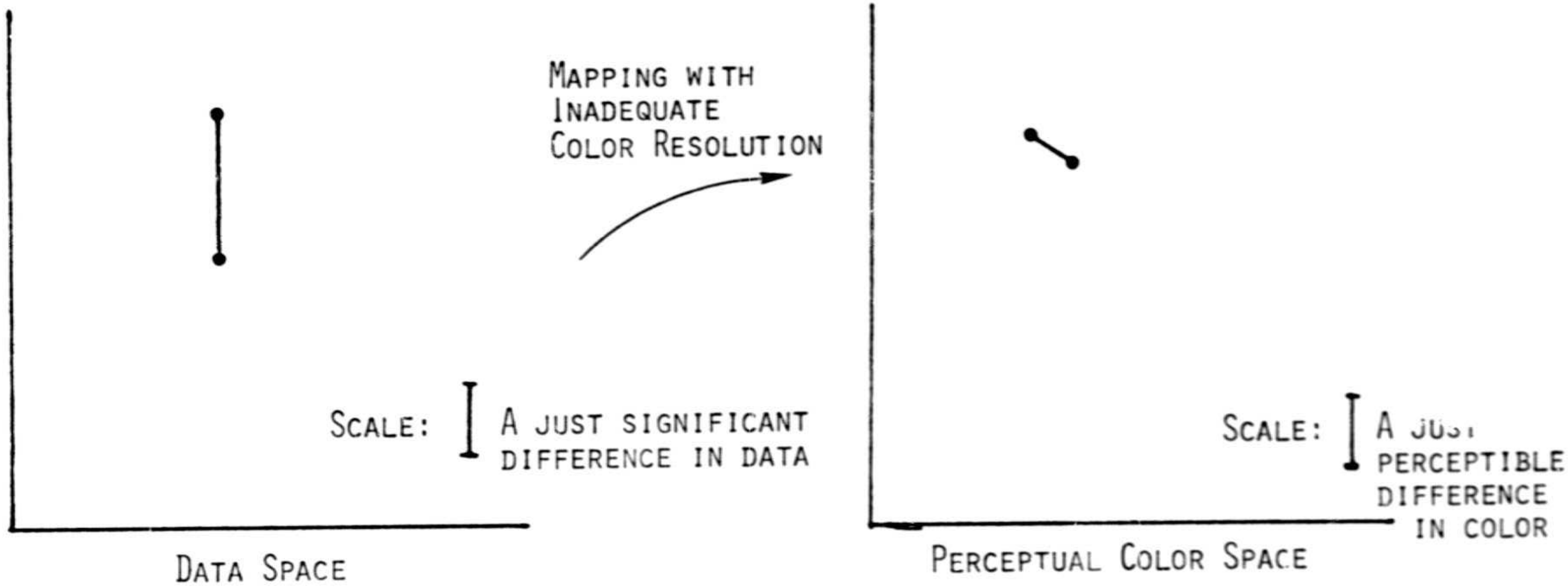
  1) USE PRINCIPLES OF COLOR PHOTOGRAPHY AND COLOR SCIENCE TO MODEL THE COLOR PRODUCTION CHARACTERISTICS OF THE PFC.

  2) APPLY EQUATIONS OF COLORMETRICS TO SPREAD THE PFC COLORS OUT INTO APPROXIMATELY UNIFORM PERCEPTUAL SPACING.

  3) MAKE IMAGERY BY MAPPING DATA LINEARLY INTO THIS UNIFORM COLOR SPACING

FIGURE 8. AVOID MASKING VARIATION

- CAPTURE DATA VARIATION WITHOUT LOSS



MOTIVATION

WHEN SIGNIFICANT DATA VARIATION IS MASKED, THE INFORMATION AVAILABLE FOR LABELING IS DIMINISHED.

FIGURE 9. APPROACH TO OBTAINING APPROPRIATE RESOLUTION

- SCALE DATA INTO UNIFORM COLOR SPACE SUCH THAT A SIGNIFICANT DIFFERENCE BECOMES A PERCEPTIBLE DIFFERENCE

  1) STATE VALUE OF SIGNIFICANT DIFFERENCE IN EACH DATA DIMENSION IN SCANNER COUNTS.

  2) FIND SIZE OF PERCEPTIBLE COLOR DIFFERENCE IN UNIFORM COLOR SPACE UNITS.

  3) DIVIDE TO ESTABLISH DESIRED RESOLUTION IN UCS UNITS/COUNT.

  4) REQUIRE THAT MAP OF DATA INTO COLOR SPACE MEET THE DESIRED COLOR RESOLUTION.

FIGURE 10. MATCH DATA DIMENSION WITH PERCEPTUAL DIMENSION

- MATCH INDEPENDENT DIMENSIONS OF DATA VARIATION TO COLOR VARIATION

INDEPENDENT DATA FEATURES	INDEPENDENT COLOR ATTRIBUTES
BRIGHTNESS FEATURE	BRIGHTNESS
GREENNESS FEATURE	CHROMATICITY
YELLOWNESS FEATURE	

MOTIVATION

- EASY INTERPRETATION OF COLOR VARIATION

ERIM

FIGURE 11. APPROACH TO MATCHING DIMENSIONS OF VARIATION

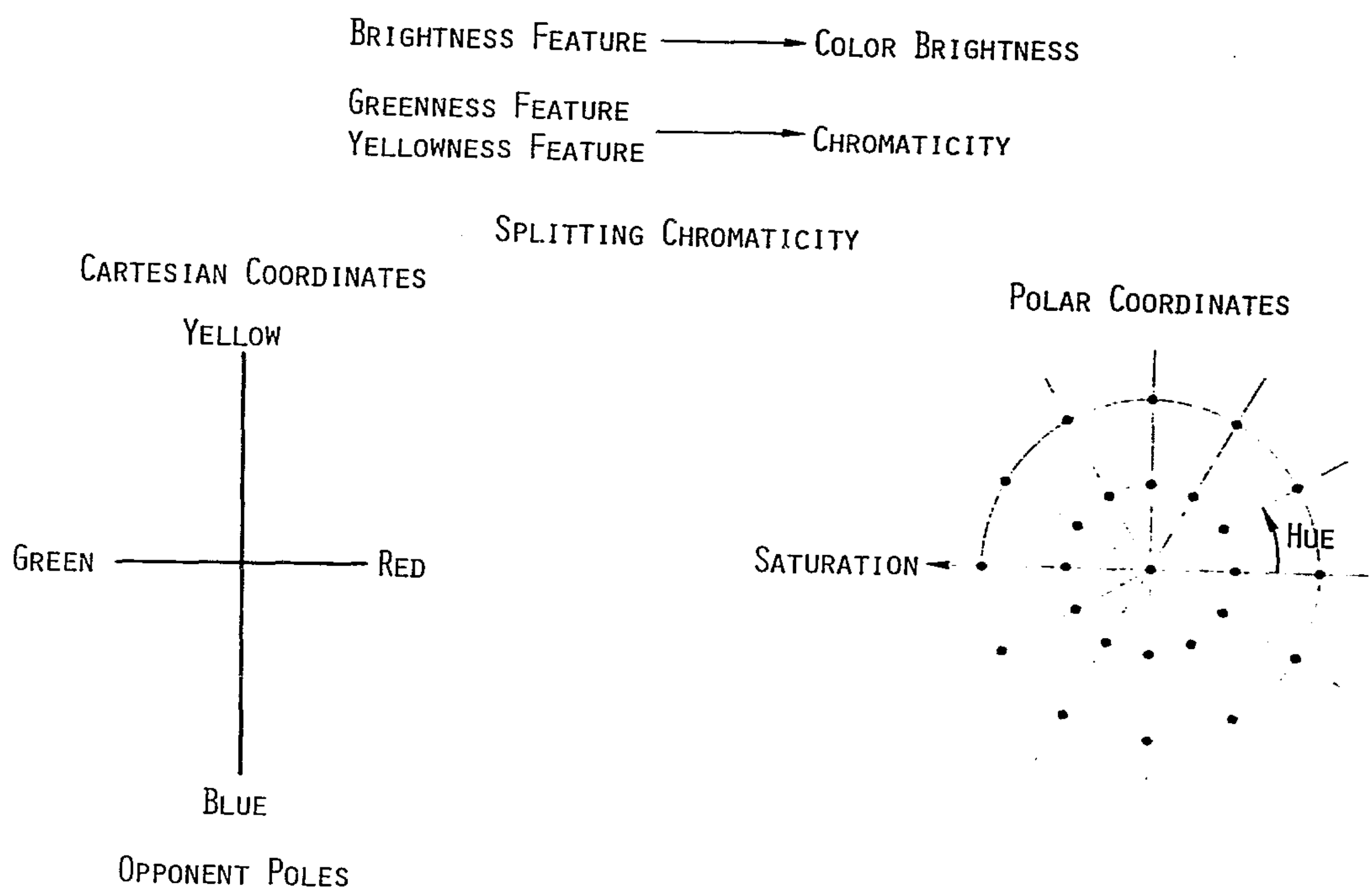
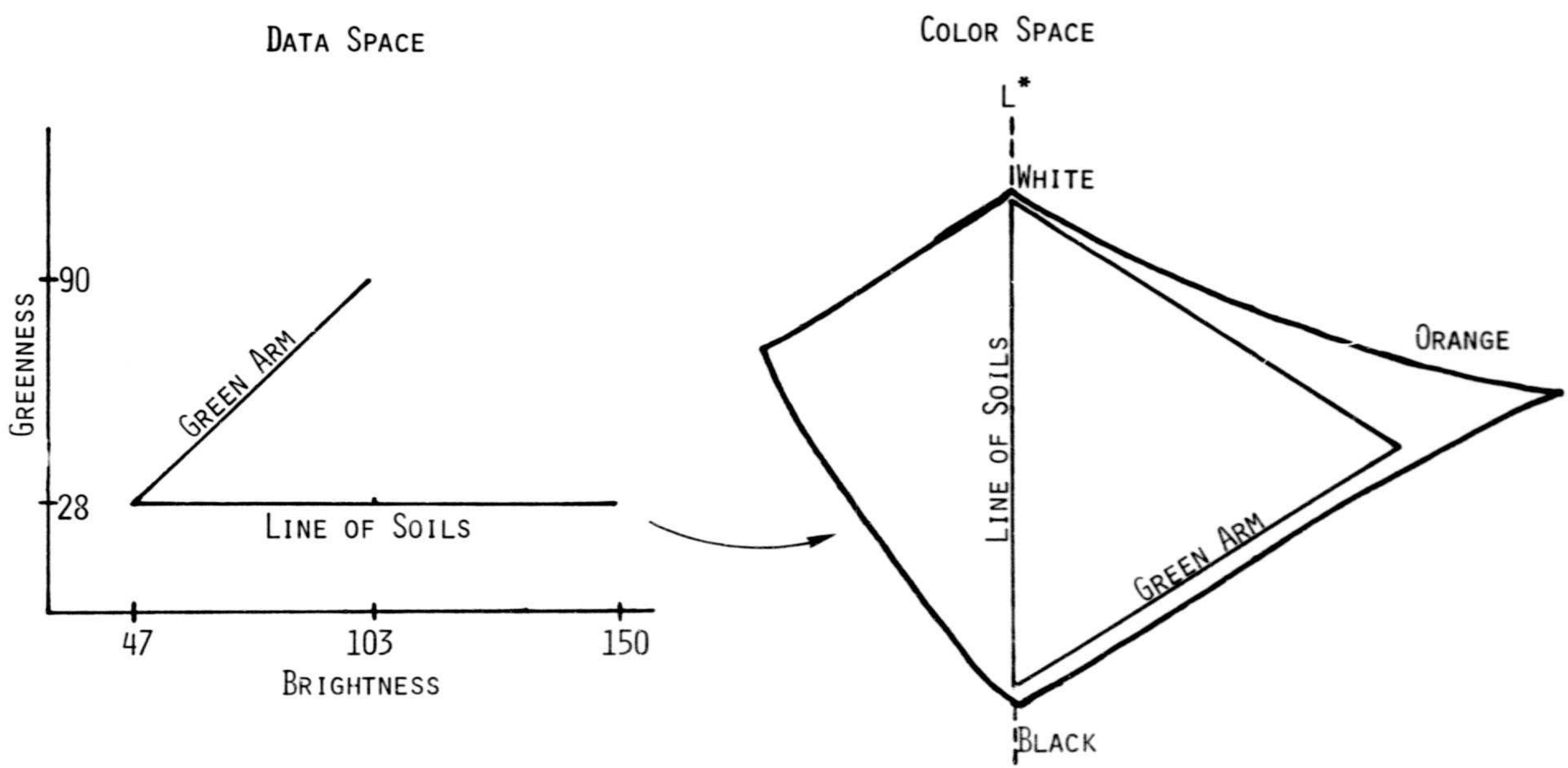


FIGURE 12. BRIGHTNESS AND GREENNESS IN THE EXPERIMENTAL MAPPING



BRIGHTNESS RESOLUTION = 0.7 UCS UNITS PER COUNT

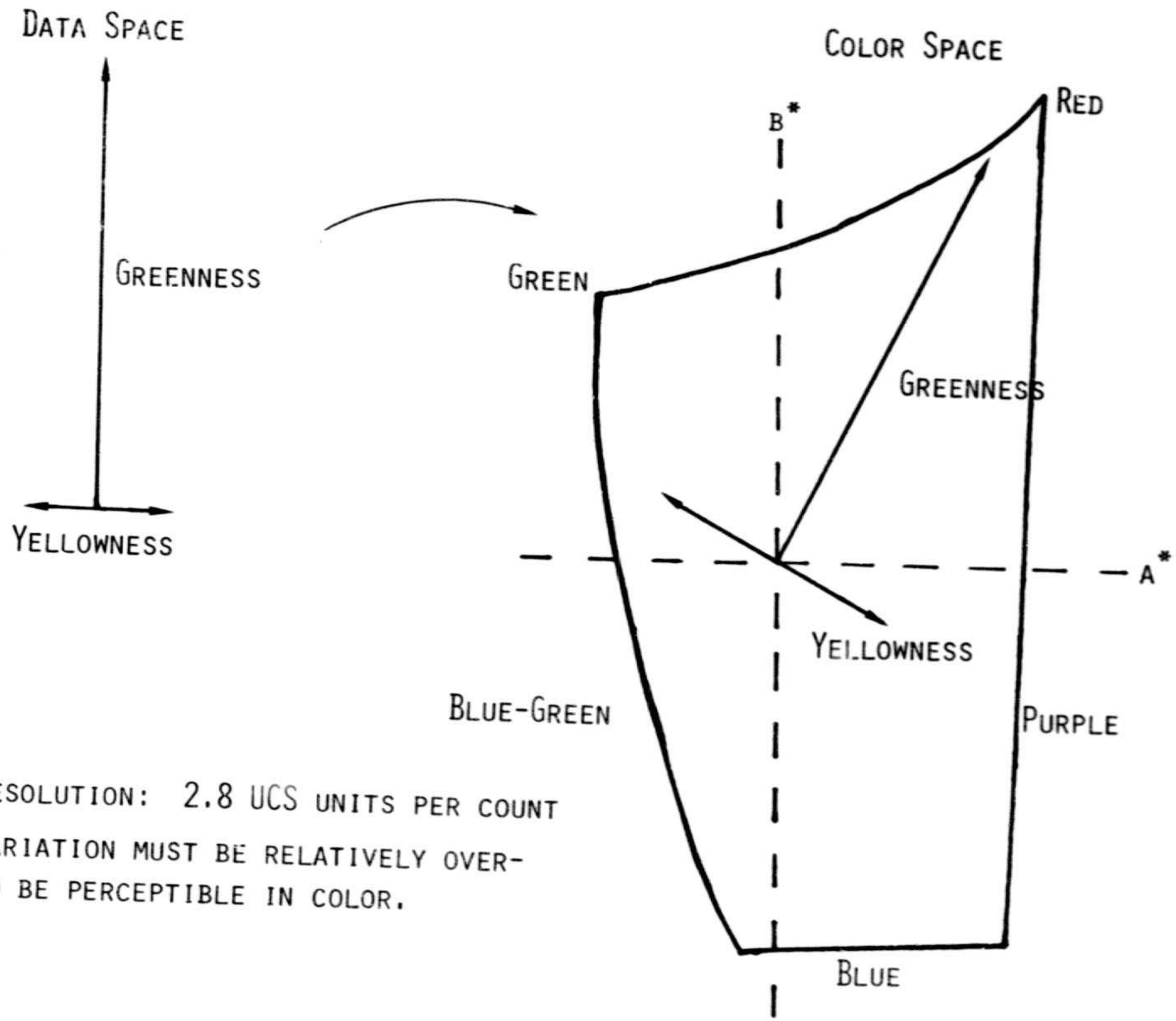
GREENNESS RESOLUTION = 1.4 UCS UNITS PER COUNT

A JUST PERCEPTIBLE COLOR DIFFERENCE IS ROUGHLY 5 UCS UNITS.

THE LESS NOISY GREENNESS DIRECTION IS PROVIDED GREATER COLOR RESOLUTION.

ERIM

FIGURE 13. GREENNESS AND YELLOWNESS IN THE EXPERIMENTAL MAPPING



YELLOWNESS RESOLUTION: 2.8 UCS UNITS PER COUNT

YELLOWNESS VARIATION MUST BE RELATIVELY OVER-ENHANCED TO BE PERCEPTIBLE IN COLOR.